**3K**

**УДК 004.89:004.93**

***Ю.В. Крак*[1], *О.В. Бармак*[1], *С.О. Романишин*[2]**
[1] Інститут кібернетики ім. В.М. Глушкова НАН України, м. Київ, Україна
[2] Хмельницький національний університет, Україна
serg.romanyshyn@gmail.com

# Узагальнені граматичні конструкції для автоматизованого перекладу з української мови на українську жестову мову

У статті розглядається проблема побудови системи автоматизованого перекладу з української мови на українську жестову мову. Пропонується інформаційна технологія, що дозволяє побудувати таку систему. Основна ідея полягає у створенні механізму узагальнених граматичних конструкцій простих речень української та жестової мов. Отримані узагальнені конструкції пов'язані між собою зв'язками, які дозволяють робити переклад, не гублячи при цьому сенс тексту. Проведено експериментальне дослідження на множині слів та жестів в об'ємі початкових класів спеціальних шкіл для дітей з вадами слуху.

## Вступ

Створення нових сучасних комп'ютерних систем навчання та комунікації для людей з вадами слуху, за допомогою яких можна було б вирішити актуальні завдання сучасності, зокрема, реалізацію інклюзивної освіти та загальне вивчення жестової мови, є важливою і актуальною проблемою. Для вирішення цієї проблеми запропонована концепція інформаційної технології невербального спілкування людей з вадами слуху [1]. Комплексна інформаційна технологія включає в себе функціональність за синтезом: рухів жестової мови, дактильної абетки, міміки промовляння на просторовій моделі людини та ін.

Одним з напрямків концепції є створення автоматизованої системи перекладу з української мови на українську жестову мову. **Метою даної роботи** є створення засобів аналізу й синтезу граматичних конструкцій речень цих мов для побудови такої системи.

## Система жестового спілкування глухих

Система жестового спілкування глухих має два різновиди: розмовну жестову мову і калькуючу жестову мову. Різновиди відрізняються за лінгвістичною структурою, функціональним призначенням та умовами їх опанування [2]. Калькуюча жестова мова – це жести, які супроводжують усне мовлення того, хто говорить. Жести, що виступають як еквіваленти слів (зазвичай промовлених глухими без голосу тільки губами), слідують в тому ж порядку, в якому розташовані слова звичайного речення. Не маючи власної граматики, калькуюча жестова мова будується на основі граматики словесної мови. Калькуюча жестова мова зазвичай використовується глухими нашої країни в офіційній обстановці – на зборах, конференціях тощо.

Розмовна жестова мова – це спілкування за допомогою засобів розмовної жестової мови. У неї своєрідна граматика. Наприклад, якщо фраза «У лівому нижньому кутку кімнати стоїть стілець, за стільцем торшер» підкріплюється в калькуючій мові декількома жестами, що позначають окремі поняття (ЛІВИЙ, НИЗЬКО, КУТ, КІМНАТА, СТОЯТИ,

стільці, торшер) і буквами алфавіту (ЗА СТІЛЕЦЬ-о-м), то в розмовній жестовій мові ця фраза демонструється таким чином: ліва рука показує жест, що означає поняття СТІЛЕЦЬ, права рука – торшер, причому обидва жести виконуються одночасно, але лівіше і ближче до розповідача – відображаючи тим самим задане просторове положення предметів.

Надалі розглядатимемо розмовну жестову мову. Специфіка лексики у ній зумовлена тим, що розмовна жестова мова використовується в невимушеній, неофіційній обстановці, коли ведуться розмови про події повсякденного життя. Тому деякі поняття у мові відсутні, наприклад спеціальні позначення, що застосовуються тільки в процесі навчання. Коли виникає потреба у вираженні подібних понять, глухі люди, які досягли певного рівня освіти та розвитку, використовують калькуючу жестову мову або словесну мову. Однак це зовсім не говорить про неповноту або примітивність розмовної жестової мови, а тільки підкреслює специфіку її функціонального призначення. Ця специфіка зумовлює і відсутність у лексиці мови спеціалізованих позначень, завжди присутніх у ситуації розмови, наприклад, позначень голови, носа, руки тощо. Вони завжди виражаються вказуванням на голову, ніс, руку тощо. Треба зауважити, що вказівні жести широко використовуються в розмовному жестовому мовленні і мають досить широкий діапазон функцій. Наприклад, якщо глухий «розповідає» співрозмовнику, якого кольору його нове пальто, то він може просто вказати на перехожого, показавши тим самим, що колір його пальто такий же, як і колір пальто цього перехожого. Особливості лексики розмовної жестової мови можуть бути пов’язані і з кінетикою. Наприклад, жести, що виражають значення «підніматися» (в гору) і «спускатися» (з гори), різняться лише напрямом руху: знизу вгору і згори вниз, а значення «дивитися» і «оглядати» – тільки якістю руху: рівномірний рух і круговий рух.

Лексика розмовної жестової мови ще недостатньо вивчена, особливо ідіоматика, фразеологія, морфологія. Автоматичне приписування жестовій мові усіх форм словесної та письмової мови є невірним, оскільки неможливо вставити розмовну жестову мову глухих у традиційні моделі лінгвістики. Для опису фактів морфології розмовної жестової мови найбільше підходить опис за принципом: від значення – до форми. Значення «множинності», «часу», «приналежності», «закінчення дії» передаються у жестовому мовленні своїми способами. Залежно від значення змінюється і спосіб виконання жесту (якість руху). Наприклад, жести у значенні «доручати», «купувати», «вирішувати» виконуються повільно і кожен жест повторюється кілька разів, а жести в значенні «доручити», «купити», «вирішити» – виконуються різко і одноразово. Зміна способу виконання жесту (інфлексія) – найпоширеніший спосіб подання суб’єктно-об’єктних відносин у розмовному жестовому мовленні. Рух від себе показує роль об’єкта (дивлюся, допомагаю), а рух до себе – роль суб’єкта (дивись на мене, допомагай мені).

Отже, узагальнюючи вищевикладене, можна зробити висновок, що розмовна жестова мова – це самобутня мова, що дозволяє висловити будь-які смисли і відносини між смислами. Тому вона успішно вирішує проблему спілкування глухих у невимушеній, неофіційній обстановці.

Відзначимо, що у звичайній мові основою є слово, тоді як у жестовій мові жест є основною семантичною одиницею зі складною структурою. Жест складається з таких компонентів, як конфігурація, місце жесту і характеристики руху руки.

Жест у жестовій мові глухих є рухом руки (двох рук) і пальців. Руки, рухаючись, торкаються певним чином обличчя, голови, тулуба, які пасивно беруть участь у жесті. Міміка обличчя (рухи частин обличчя – брів, губ, роту) є важливою складовою у жестовій мові. У процесі жестового спілкування вона відіграє значно більшу роль, ніж в процесі усного мовлення.

Для вираження певних відношень у жестовій мові використовуються два порядки побудови речень: лінійний та одночасний. При лінійному порядку жестові одиниці слідують одна за одною, а при одночасному – одна рука відтворює один жест, а інша – інший жест в один і той же відлік часу.

Розглянемо синтаксичні особливості жестової мови для простих речень. Підмет і присудок у реченнях пов'язані предикативним зв'язком. Під простими реченнями розумітимемо речення з одним предикативним зв'язком. У жестовій мові прості речення поділяються на розповідні, питальні та спонукальні.

Розповідне жестове речення містить у собі повідомлення про якийсь факт, розповідь про щось. Воно поділяється на стверджувальне та заперечне речення. Для побудови стверджувального розповідного речення застосовуються наступні правила:

1. При побудові розповідного речення лінійного порядку першою ставиться жестова одиниця, що позначає об'єкт, який відповідає на питання хто? що? (аналог в усній мові – підмет). Слідом відтворюється жест, який вказує дію (присудок). Наприклад:

Я купила ⟶ Я / КУПИТИ / ВЖЕ

2. При розширенні жестового речення жестом, який означає якість предмета (прикметник), першим відтворюється жест, що позначає об'єкт, а потім жест, який позначає якість. У випадку декількох жестів, які позначають об'єкт, жест, який позначає якість, відтворюється після відповідного жесту, який позначає об'єкт:

Смачний пиріг на тарілці ⟶ ТАРІЛКА / ПИРІГ / СМАЧНИЙ

3. При побудові розповідного стверджувального речення із застосуванням жесту, що позначає час, місце, причину тощо (прислівник), першим відтворюється жест, який позначає об'єкт, наступним – жест, який уточнює обставину для зазначеного об'єкта. Наприклад:

На тарілці мало пирогів ⟶ ТАРІЛКА / ПИРІГ / МАЛО

4. Якщо в усній мові логічним наголосом підкреслюється базове слово, то у жестовій мові базовий жест для чіткої візуалізації застосовується останнім. Наприклад, у жестовій конструкції: СЬОГОДНІ / БУДЕ / БАНКЕТ співрозмовник розуміє, що *сьогодні відбудеться не щось інше (екскурсія, конференція), а саме банкет*. Якщо подане речення завершити жестом СЬОГОДНІ, то зміст буде наступним: *банкет буде не завтра, а саме сьогодні*.

Побудова окличного речення в жестовій мові аналогічна до побудови простого розповідного речення. При візуальному сприйманні та відтворенні окличного жестового речення підвищується його експресивність. Змінюється характеристика рухів рук, які стають швидшими у момент відтворення жестових одиниць, дещо різкими залежно від змісту, а також збільшується амплітуда рухів рук. Важливим є емоційний вираз обличчя, який в різних окличних реченнях містить спектр почуттів від гніву до радощів. Третім чинником є зміна верхньої частини тулуба: наприклад, нахил вперед голови і плечового пояса.

У жестовій мові заперечні жести НІ, НЕ БУДЕ, НІКОЛИ, НЕ ХОЧУ тощо, які є основою розповідних жестових конструкцій, застосовуються лише після жестів, які вони заперечують. Наприклад, якщо треба передати сенс: «Я не хочу тебе бачити» послідовним відтворенням жестів: Я НЕ ХОЧУ ТЕБЕ БАЧИТИ, то нечуючі можуть зрозуміти це так: Я не хочу. ТИ це бачиш. А тому в даному реченні потрібно застосувати наступну послідовність: Я БАЧИТИ ТЕБЕ НЕ ХОЧУ. Це перше мовне правило побудови заперечних речень.

Друге правило полягає у тому, що слова усного мовлення, які несуть в собі заперечення, в жестовій мові, по можливості, замінюються не відповідними, а іншими жестами. Наприклад, не дорогий – ДЕШЕВИЙ, не весело – СУМНО або «не» ставиться після певного слова, наприклад, «Я не працюю» передається так: Я / ПРАЦЮЮ / НЕ.

Питальні речення містять у собі якесь запитання, для отримання інформації. Вони поділяються на власне питальні (містять пряме питання, що вимагає обов'язкової відповіді) та риторичні (що не вимагають відповіді).

В жестовій мові існує два правила для послідовної побудови питального речення:

1. Жест, який містить у собі запитання, завжди ставиться у кінці речення. Наприклад:

Скільки буде двічі два? → ДВА МНОЖИТИ ДВА БУДЕ СКІЛЬКИ

Чи пробувала ти цей салат? → ТИ САЛАТ ЦЕЙ ПРОБУВАТИ ВЖЕ

Коли питальне речення містить декілька жестових одиниць, які позначають предмет, то питальні жести: ХТО, КОЛИ, ДЕ, КУДИ, ЗВІДКИ, ЯК, тощо – можуть повторюватись для підсилення питальної конструкції, наприклад:

Як вони можуть перелізти паркан? → ЯК ВОНИ МОЖЕ ПАРКАН ПЕРЕЛІЗТИ ЯК

2. Вираз обличчя, «виразність» жестів, які несуть питальне навантаження, довші паузи.

Спонукальне речення виражає волевиявлення мовця, яке потребує виконання: наказ, прохання, пораду, застереження, протест.

# Автоматизований переклад

Автоматизований переклад (Computer-Aided Translation) - це переклад текстів на комп'ютері з використанням комп'ютерних технологій [3]. Це широке поняття, що може включати в себе різні засоби та інструменти: перевірку правопису та розділових знаків, пошук перекладу у раніше перекладених текстах, термінологічні бази даних, системи, що дозволяють користувачу вносити правки в сумнівних випадках. Від машинного перекладу він відрізняється тим, що процес перекладу виконується людиною, комп'ютер лише допомагає їй виконати переклад або за менший час, або з кращою якістю.

Автоматизований переклад передбачає такі форми взаємодії з користувачем: 1) частково автоматизований переклад (використання людиною-перекладачем комп'ютерних словників); 2) системи з розподілом праці, в яких комп'ютер навчений перекладати тільки фрази, структура яких жорстко задана (але робить це так, щоб виправляти за ним не було потреби), а фрази з іншою структурою віддає на переклад людині.

Так званий «прямий» переклад, що застосовується у багатьох системах, передбачає перетворення тексту із однієї форми подання (вхідна мова) в іншу (вихідна мова) за словами, не додаючи та не вилучаючи нічого із тексту. Тобто кожне слово перекладається окремо, таким чином, отриманий в такому випадку переклад є сумою перекладів окремих слів. Даний тип перекладу доволі простий, але має багато недоліків, оскільки в мовах дуже мало слів, що перекладаються однозначно. У більшості випадків одне слово має декілька значень і навпаки. Вибір кращого варіанта залежить від контексту, в якому використовується слово, що перекладається. Тобто мова йде про передачу сенсу тексту. Деякі слова або словосполучення при цьому можуть перекладатися з адаптацією до мови перекладу або ж зовсім опускатися. Головне – точно передати сенс висловлювання іншою мовою. Такий тип перекладу має на увазі, що його здійснюють групи перекладачів з філологічною освітою на рідній мові.

Виходячи зі сказаного вище, сформульована наступна постановка задачі: запропонувати комп'ютерну технологію, що реалізуватиме систему автоматизованого перекладу речень українською мовою на відповідні конструкції-речення української жестової мови.

Пропонується розв'язок поставленої задачі з наступними обмеженнями: 1) до уваги беруться тільки прості речення; 2) сенс жестової мови обмежується фіксованим переліком тем і ситуацій; 3) система перекладає (без наступних виправлень) тільки навчені жорстко задані структури фраз та речень і не спотворює сенсу; 4) можливе постійне розширення (навчання) переліку структур перекладу.

# Система автоматизованого перекладу з української мови на українську жестову мову

Система автоматизованого перекладу з української мови на жестову мову передбачає створення множинної моделі української мови [4-7] та реалізації її для можливості моделювання відповідних пов'язаних граматичних конструкцій української та жестової мов. В основі системи лежить граматична конструкція простих речень з відповідним групуванням їх за темами, що виникають у повсякденному житті. Під простим реченням розуміється речення, у якому є тільки одна граматична основа (предикативний центр).

Для досягнення мети перекладу необхідно запропонувати граматичний словник української мови. Українська мова відноситься до флективної мови – мови синтетичного типу з домінуванням словозмін за допомогою флексій – формантів, що поєднують відразу декілька значень. Флективна будова мови протистоїть аглютинативній, у якій кожен формант несе тільки одне значення. За основу для розробки граматичного словника візьмемо теорію лексикографічних систем [4-7].

Побудова граматичного словника флективної мови визначається наявністю нормальної моделі словозміни, що означає встановлення та формалізацію лінгвістичних критеріїв, згідно з якими вся множина слів мови розбивається на певні підмножини, взаємний перетин яких є порожнім, а всередині кожної з них словозміна відбувається за однаковими правилами. Підмножини слів з такими властивостями називаються словозмінними парадигматичними типами.

Під парадигматичним типом розумітимемо групу лексем, словозмінна парадигма яких характеризується однаковою кількістю граматичних форм, усередині якої словозміна відбувається за тим самим (єдиним) правилом. Для української мови, яка є мовою аналітико-синтетичного типу, це означає, що, по-перше, слова, які належать до одного парадигматичного класу, мають однакові флексії у відповідних граматичних значеннях та однаковий характер чергування в основі і, по-друге, відповідні аналітичні форми будуються за однаковими моделями їх утворення.

Українській мові притаманні наступні граматичні категорії, що визначають словозміну: $s$ – число (однина($s_1$), множина($s_2$)), $g$ – рід (чоловічий ($g_1$), жіночий ($g_2$), середній ($g_3$), вищий ступінь ($g_4$), чоловічий або середній рід для істоти ($g_5$), жіночий або середній рід для істоти ($g_6$), чоловічий або жіночий рід для істоти ($g_7$), незмінювана словникова одиниця ($g_8$), найвищий ступінь ($g_9$), $i$ – узагальнена категорія, що включає відмінок, особу, час, стан, спосіб (називний ($i_1$), родовий ($i_2$), давальний ($i_3$), знахідний ($i_4$), орудний ($i_5$), місцевий ($i_6$), кличний ($i_7$), інфінітив ($i_8$), 1-а особа, наказовий спосіб ($i_9$), 2-а особа, наказовий спосіб ($i_{10}$), 3-я особа, наказовий спосіб ($i_{11}$), 1-а особа, майбутній час ($i_{12}$), 2-а особа, майбутній час ($i_{13}$), 3-я особа, майбутній час ($i_{14}$), 1-а особа, теперішній (майбутній) час ($i_{15}$), 2-а особа, теперішній (майбутній) час ($i_{16}$), 3-я особа, теперішній (майбутній) час ($i_{17}$), 1-а особа, теперішній час ($i_{18}$), 2-а особа, теперішній час ($i_{19}$), 3-я особа, теперішній час ($i_{20}$), активний дієприкметник, теперішній час ($i_{21}$), дієприслівник, теперішній час ($i_{22}$), минулий час ($i_{23}$), активний дієприкметник, минулий час ($i_{24}$), пасивний дієприкметник, минулий час ($i_{25}$), безособова форма, минулий час ($i_{26}$), дієприслівник, минулий час ($i_{27}$), незмінювана одиниця ($i_{28}$)).

Кожне слово української мови віднесемо до наступних класів (частин мови) ( $p$ ): іменники ( $p_1$ ), числівники порядкові ( $p_2$ ), числівники кількісні ( $p_3$ ), числівники ( $p_4$ ), числівники типу «два» ( $p_5$ ), дієслова недоконаного і доконаного виду ( $p_6$ ), дієслова доконаного виду ( $p_7$ ), дієслова недоконаного виду ( $p_8$ ), дієприкметники ( $p_9$ ), прикметники ( $p_{10}$ ), займенники ( $p_{11}$ ), займенники-прикметники ( $p_{12}$ ), прислівники ( $p_{13}$ ), вигуки ( $p_{14}$ ), сполучники ( $p_{15}$ ), частки ( $p_{16}$ ), прийменники ( $p_{17}$ ), присудкові слова ( $p_{18}$ ), вставні слова ( $p_{19}$ ), абревіатури ( $p_{20}$ ), сполуки ( $p_{21}$ ), дієприслівники ( $p_{22}$ ), займенники з часткою ( $p_{23}$ ), сполучники і частки ( $p_{24}$ ), прислівники і частки ( $p_{25}$ ), числівники з прийменником ( $p_{26}$ ), іменники з прийменником ( $p_{27}$ ), дієприслівники з часткою ( $p_{28}$ ), займенники з прийменником ( $p_{29}$ ), дієприслівники з часткою ( $p_{30}$ ).

Для української мови введемо наступні парадигматичні типи (табл. 1).

Таблиця 1 – Парадигматичні типи української мови

| Парадигматичні типи | Граматичні класи | Граматичні категорії, що визначають словозміну | Кількість граматичних значень |
|---|---|---|---|
| $T^1$ | $p_1$ | $s*(i_1,\ldots,i_7)$ | 14 |
| $T^2$ | $p_2 \cup p_4 \cup p_9 \cup p_{10} \cup p_{12}$ | $(g_1,g_2,g_3)*((s_1,s_2)*(i_1,\ldots,i_7))$ | 24 |
| $T^3$ | $p_3 \cup p_{11}$ , | $(i_1,\ldots,i_6)$ | 6 |
| $T^4$ | $p_5$ | $(g_1,g_2)*(i_1,\ldots,i_6)$ | 12 |
| $T^5$ | $p_6 \cup p_8$ | $i_8,(i_{10}*s_1),(i_9*s_2),(i_{10}*s_2),(i_{12}*s_1)$ $((i_{12},\cdots,i_{17})*s),i_{21},i_{22},$ $i_{23}*(g_1,g_2,g_3)*s_1),$ $(i_{23}*s_2),i_{24},i_{25},i_{26},i_{27}$ | 26 |
| $T^6$ | $p_7$ | $i_8,(i_{10}*s_1),(i_9,s_2),(i_{10},s_2),$ $((i_{12},i_{13},i_{14})*s),$ $(i_{23}*(g_1,g_2,g_3)*s_1),$ $(i_{23}*s_2),i_{24},i_{25},i_{26},i_{27}$ | 18 |
| $T^0$ | $\bigcup_{j=13}^{30} p_j$ | $i_{28}$ | 1 |

Виходячи з того, що українська мова є флективною мовою (тобто граматичні значення передаються флексіями), слова мови моделюються у вигляді комбінації незмінної та змінної складових:

$$x = c(x)\,\&\,f(x), \tag{1}$$

де $c(x)$ – частина лексеми *x*, яка в процесі словозміни залишається незмінною (квазіоснова), $f(x)$ – її змінна складова (квазіфлексія), & – конкатенація.

Реалізація такої моделі зберігання слова прибирає надлишковість, маючи на увазі, що в українській мові більше мільйона слів (у всіх словозмінах) та тільки біля ста тисяч базових слів – інфінітивів. Тобто замість повного тексту слова у таблиці зберігається

номер слова-інфінітива із відповідної множини інфінітивів, номер позиції у слові, до якої слово-інфінітив незмінне, та номер флексії (змінної частини слова) із відповідної множини. Тобто слова української мови подаються у вигляді:

$$W = \{W_i : W_i = \{I_{i_1} \in I, F_{i_2} \in F, k, g \in G, s \in S, in \in In\}\}, \quad (2)$$

де $W_i$ – параметри слова української мови ($i = 0, \cdots, N-1$, $N$ – кількість слів у словнику); $F$ – множина всіх можливих закінчень слів (флексій); $k$ – номер позиції у слові-інфінітиві, з якої починає конкатенуватися флексія (можливі випадки, коли $k = 0$ – для словоформи, яка повністю відмінна від інфінітива); $G = \{g_1, \ldots, g_9\}$; $S = \{s_1, s_2\}$; $In = \{i_1, \ldots, i_{28}\}$ $I$ – множина слів-інфінітивів української мови (для дієслів – інфінітиви, для іменників – слова у називному відмінку, однині тощо):

$$I = \{I_i : I_i = \{word\inf, p \in P\}\}, \quad (3)$$

де $P \in \{p_1, \ldots, p_{30}\}$; $word\inf$ – слово-інфінітив.

Формування множин $W, F, I$ проходить наступним чином (для всіх слів-інфінітивів мови):

- *Вхід*: слово-інфінітив та усі його словоформи (граматичні значення)
- $k =$ Довжина(загальна частина для слова-інфінітива та всіх його словоформ)
- Для всіх словоформ визначаємо закінчення (відмінна частина) та додаємо його до множини $F$ (якщо такого закінчення там немає)
- Додаємо слово-інфінітив до множини $I$
- Формуємо множину $W$ (2)

Прямою задачею для такої моделі (2) є проста задача відтворення повного тексту слова за трьома параметрами $I_{i_1}, F_{i_2}, k$, яка реалізується відповідним оператором ($H$):

$$H = \begin{cases} Left(I_{i_1}, k-1) \& F_{i_2}, & k \neq 0 \\ Left(I_{i_1}, 0) \& F_{i_2}, & k = 0 \end{cases}, \quad (4)$$

де $Left(Word, Length)$ – функція отримання перших $Length$ символів із слова $Word$; & – конкатенація.

Деяка складність виникає для оберненої задачі ($H^{-1}$): для існуючого слова знайти три параметри, які його однозначно визначають. Реалізація оператора (4) для бази даних у вигляді недетерміністської функції призводить до того, що неможливо ефективно проіндексувати. Пошук по неіндексованому полю – досить затратний за часом. Для вирішення цієї проблеми запропонований алгоритм, сенс якого полягає у тому, що вхідне слово розбивається на можливі комбінації незмінної та змінної складових і пошук проводиться по проіндексованих полях таблиць-множин інфінітивів ($I$) та флексій ($F$):

$H^{-1} =$ (5)

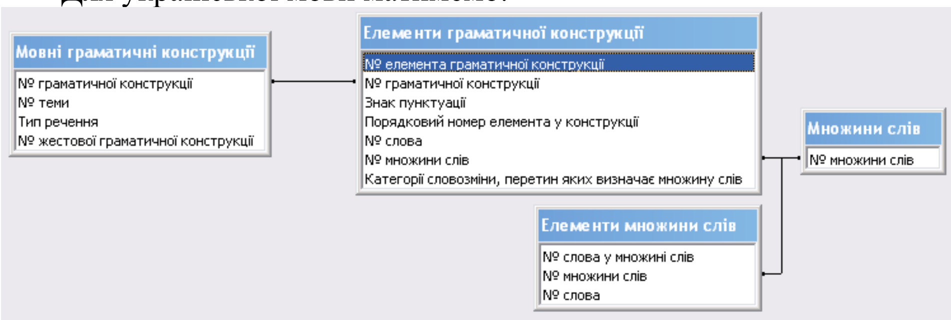
```
i=0
l=Довжина(Word)
Повторювати поки i<l
     Head=LEFT(Word,i)
     Tail=RIGHT(Word,l-i)
     Якщо існує інфінітив I_i1 ∈ I, що починається на Head та
          існує флексія F_i2 ∈ F, що дорівнює Tail, то
          визначаємо (I_i1, F_i2, k)
     Інакше
          i=i+1
Кінець циклу
```

Отже, до розгляду беремо тільки прості речення, тобто вважатимемо, що довільне складне речення можна подати як декомпозицію простих речень, причому обмежимося наступними типами простих речень: розповідні (стверджувальні, окличні та заперечні), питальні та спонукальні.

Речення в моделі даних об’єднуватимуться у структури, отримані шляхом узагальнення. Структури речень міститимуть замість слів їх послідовності, кожна з яких може використовуватись при побудові речення. Послідовності можуть мати як окремі слова, так і різноманітні множини. Наприклад, речення «він йде», «вона йде», «час йде» об’єднуються в одну структуру «{він, вона, час, хто, …} йде». Дані послідовності можуть змінюватись та доповнюватись у процесі перекладу. Аналогічні структури будуватимуться і для жестових речень.

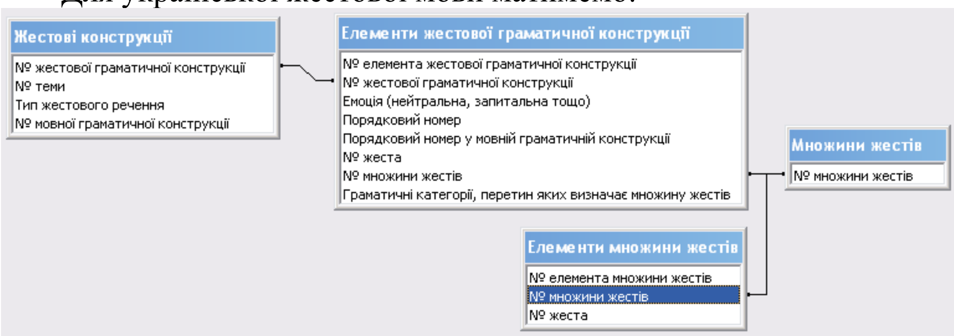
Формальну модель узагальнених граматичних конструкцій (маючи на увазі їх майбутню реалізацію у реляційній базі даних) подамо у вигляді наступних таблиць та зв’язків між ними.

Для української мови матимемо:




(6)

Для української жестової мови матимемо:




(7)

Для автоматизованого перекладу тексту з української мови на українську жестову мову використовується наступна алгоритмічна схема:

1. Речення поступає на вхід, виконується пошук структури речення, в яку входять кожне зі слів. Йде запит до статистичної бази: чи було таке речення оброблене раніше. Статистична база містить інформацію про частоту повторів перекладу даного речення, структуру даного речення, тематику даного речення, наявність змін у перекладі при обробленні його користувачем.

2. Якщо речення міститься у статистичній базі, воно обробляється і видається результат перекладу. В іншому випадку проводиться аналіз речення. Структура речення порівнюється з структурою оброблених раніше речень. Якщо структури речень частково збігаються, робиться спроба прогнозу на основі варіанта, який найчастіше використовувався. Наприклад, якщо на вхід отримано речення і таке речення раніше не оброблялось, відбувається аналіз структури речень, в яких тематика збігається з вхідним.

3. Якщо аналогічної структури не знайдено, відбувається пошук у всіх інших тематичних розділах. Коли знайдено частину речення, що збігається з вхідним, відбувається спроба прогнозу перекладу. Наприклад, якщо на вхід дано речення

«дайте мені їсти», а в статистичній базі знайдено речення «дайте мені хліб», то для перекладу «дайте мені» використовується вже існуюча структура з бази, а для «їсти» виконується переклад з використанням іншої структури. Під час аналізу структури речення враховується його вид (розповідне, окличне розповідне, заперечне, питальне, спонукальне) та тематика.

Отримавши результат, перекладач вибирає один з варіантів:

1. Результат задовільний. Речення позначається як вірне і добавляється у статистичну базу.

2. Результат не задовільний і правиться вручну. Після цього у статистичну базу вноситься і виправлений варіант і запропонований варіант, для забезпечення прийняття рішень надалі.

3. Результат не задовільний і вважається недопустимим. Запропонований результат вноситься в статистичну базу як недопустимий. Перекладач може ввести правильний варіант перекладу, який буде збережений в базу даних та буде використовуватись при перекладі надалі.

## Реалізація системи автоматизованого перекладу

Для тестування запропонованої технології перекладу було створене відповідне програмне забезпечення.

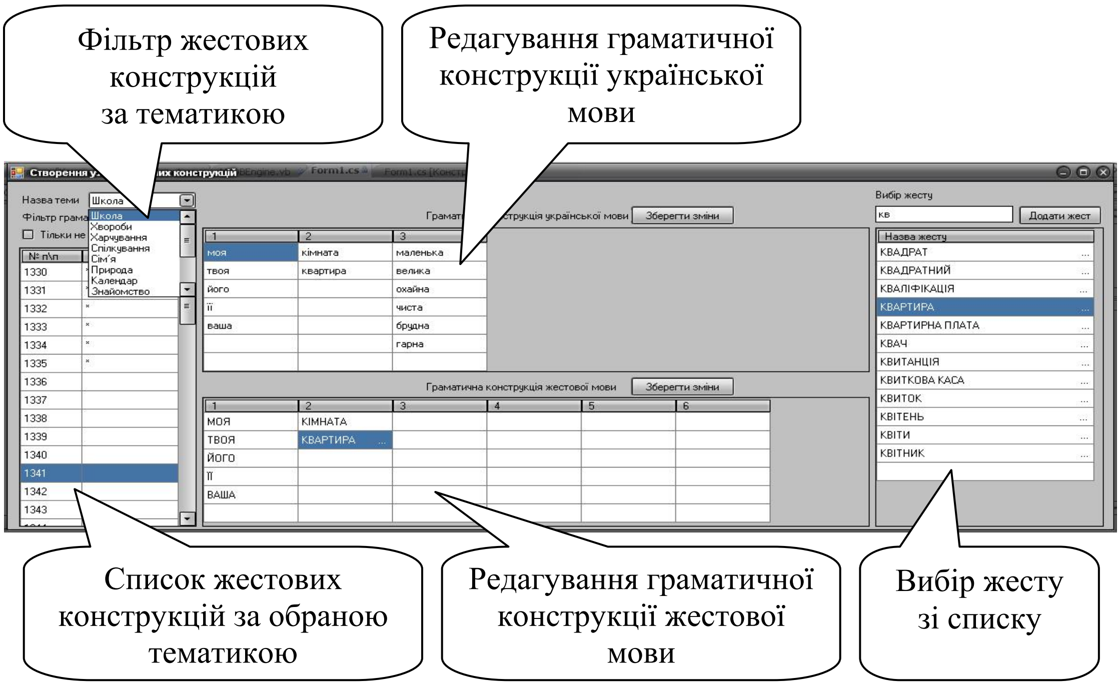



Рисунок 1 – Модуль створення узагальнюючих граматичних конструкцій

Процес створення узагальнюючих граматичних конструкцій виглядає наступним чином (рис. 1):

1) вибирається тема, для якої створюються граматичні конструкції, після чого завантажується список граматичних конструкцій, що відповідають вибраній темі;

2) після вибору граматичної конструкції з бази даних завантажуються конструкції української та української жестової мов.

Кожній узагальненій граматичній конструкції української мови ставиться у відповідність граматична конструкція жестової мови. Конструкції можуть містити як окремі слова, так і послідовності слів;

3) при створенні або зміні граматичних конструкцій жестової мови з бази даних вибираються жести, що використовуються в конструкції. Якщо потрібного жесту в базі немає, то є можливість додати новий жест. Такі жести позначаються як додані користувачем.

Автоматизований переклад у відповідному модулі здійснюється за описаною вище технологічною схемою (рис. 2).




Рисунок 2 – Модуль автоматизованого перекладу тексту з української на українську жестову мову

# Висновки

Для побудови граматичних конструкцій було взято множину речень з програми вивчення української жестової мови для спеціалізованих шкіл (для початкових класів) [8]. Було опрацьовано 2821 речень та отримано 259 узагальнених граматичних конструкцій (табл. 2). Отримані конструкції (табл. 3) покрили можливі варіанти речень по темах і з словниково-жестовим запасом для початкових класів. Навіть при такій обмеженій множині слів та жестів вдалося виділити узагальнені конструкції, які використовують не фіксовані множини слів, а ознаки граматичних категорій для групування. Наприклад, виділені конструкції:

– {цей}{( $p_1, s_1, i_1, g_1$ ) (всі іменники у називному відмінку, однині, чоловічого роду)}{( $p_{10}, s_1, i_1, g_1$ ) (всі прикметники у називному відмінку, однині, чоловічого роду)} – цей стіл мій;

– {ця}{( $p_1, s_1, i_1, g_2$ ) (всі іменники у називному відмінку, однині, жіночого роду)}{( $p_{11}, s_1, i_2, g_2$ ) (всі займенники у родовому відмінку, однині, жіночого роду)} – ця лампа моя.

Таблиця 2 – Статистика отримання узагальнюючих граматичних конструкцій

| Теми | Кількість речень | Кількість граматичних конструкцій |
|---|---|---|
| Школа | 162 | 19 |
| Здоров'я | 159 | 12 |
| Харчування | 333 | 23 |
| Спілкування | 160 | 27 |
| Природа | 311 | 30 |
| Знайомство | 354 | 43 |
| Дім | 596 | 44 |
| Анатомія | 224 | 20 |
| Побут | 522 | 41 |
| *** | 2821 | 259 |

Подальші дослідження спрямовані на доповнення системи більшою кількістю речень (за повний курс школи), що дозволить отримати переклад для множини типових речень, достатніх для звичайного спілкування.

Таблиця 3 – Приклад граматичних конструкцій для перекладу

| | |
|---|---|
| На {тарілці, столі} {ложка, виделка, пиріг, лимон} | {ТАРІЛКА, СТІЛ} {ЛОЖКА, ВИДЕЛКА, ПИРІГ, ЯБЛУКО, ЛИМОН} |
| {Я, він, брат, батько} {мені, йому, їй} купив {штани, одяг, сорочку} | {Я, ВІН, БРАТ, БАТЬКО} КУПИТИ {Я, ВІН, ВОНА} {ШТАНИ, ОДЯГ, СОРОЧКУ} ВЖЕ |
| У {дівчини, тебе, неї} {гарне, модне} плаття | {ДІВЧИНА, ТИ, ВОНА} ПЛАТТЯ {МОДНЕ, ГАРНЕ} |
| Як {тебе, його, її, вас} звати? | {ТИ, ЙОГО, ЇЇ, ВИ} ЗВАТИ *ПИТАЛЬН* |
| Це {моя, твоя, його, її} {няня, вихователька, лялька, машина} | {МОЯ, ТВОЯ, ЙОГО, ЇЇ} {НЯНЯ, ВИХОВАТЕЛЬКА, ЛЯЛЬКА, МАШИНА} |
| Що {ти, він, брат, батько} робив {вранці, вчора}? | ЩО {Я, ВІН, БРАТ, БАТЬКО} РОБИТИ {ВРАНЦІ, ВЧОРА} |
| у кімнаті є { меблі, стіл, стілець, ліжко, диван, шафа, дзеркало, килим, полиця} | У КІМНАТІ { МЕБЛІ, СТІЛ, СТІЛЕЦЬ, ЛІЖКО, ДИВАН, ШАФА, ДЗЕРКАЛО, КИЛИМ, ПОЛИЦЯ } |
| {дай, принеси} {мені, йому, їй} {чайник, тарілку, чашку, склянку, ложку, виделку, ніж, ковдру, подушку, рушник } | {ДАЙ, ПРИНЕСИ} {ЧАЙНИК, ТАРІЛКА, ЧАШКА, СКЛЯНКА, ЛОЖКА, ВИДЕЛКА, НІЖ, КОВДРА, ПОДУШКА, РУШНИК} {Я, ВІН, ВОНА} |

# Література


1. Інформаційна технологія для моделювання української мови жестів / Ю.Г. Кривонос, Ю.В. Крак, О.В. Бармак [та ін.] // Штучний інтелект. – 2009. – № 3. – С. 186-197.
2. Адамюк Н.Б. Синтаксичні особливості УЖМ: на прикладі простого речення / Н.Б. Адамюк, І.І. Чепчина // Жестова мова і сучасність : збірник наукових праць. – 2009. – № 4. – С. 170-191.
3. Bowker L. Computer-aided translation technology: a practical introduction [Електронний ресурс] / Lynne Bowker. – University of Ottawa Press, 2002. – 185 p. – Режим доступу : http://www.google.com/books?id=ly29-mc6dO0C&printsec=frontcover&hl=uk#v=onepage&q&f=false.
4. Широков В.А. Інформаційна теорія лексикографічних систем / Широков В.А. – К. : Довіра, 1998. – 331 с.
5. Широков В.А. Феноменологія лексикографічних систем / Широков В.А. – К. : Наук. думка, 2004. – 327 с.
6. Корпусна лінгвістика : монографія / [Широков В.А., Бугаков О.В., Грязнухіна Т.О. та ін.]. – К. : Довіра, 2005. – 472 с.
7. Любченко Т.П. Програмно-технологічні аспекти створення граматичних лексикографічних систем / Т.П. Любченко // Проблеми програмування. – 2007. – № 3. – С. 61-75.
8. Програма-комплекс «Українська жестова мова» [Електронний ресурс]. – Режим доступу : http://www.mon.gov.ua/education/average/programs_gluh.


# Literatura


1. Krivonos Ju.G. Shtuchnij intelekt. № 3. 2009. S. 186-197
2. Adamjuk N. B. Zhestova mova i suchasnist'. Zbirnik naukovih prac'. №4. 2009. S 170-191
3. Bowker L. Computer-aided translation technology: a practical introduction. University of Ottawa Press, 2002. 185 p. http://www.google.com/books?id=ly29mc6dO0C&printsec=frontcover&hl=uk#v=onepage&q&f=false.
4. Shirokov V.A. Informacijna teorija leksykografichnih system. K. : Dovira. 1998. 331 s.
5. Shirokov V.A. Fenomenologija leksikografichnih sistem. K. : Nauk. Dumka. 2004. 327 s.
6. Shirokov V.A. Korpusna lingvistika: Monografija. K. : Dovira. 2005. 472 s.
7. T.P. Ljubchenko T.P. // Problemi programuvannja. № 3. 2007. S. 61-75
8. 8.Programa-kompleks “UKRAINS'KA ZHESTOVA MOVA”
9. http://www.mon.gov.ua/education/average/programs_gluh


***Y.V. Krak, O.V. Barmak, S.O. Romanyshyn***

**The System of Automated Translation from Ukrainian into Sign Language**

The article is devoted to the problem of construction of automated translation from Ukrainian into sign language. An information technology that allows to build such a system is proposed. The basic idea is to create a mechanism of generalized grammatical constructions of simple sentences of Ukrainian and sign language. These result generalized structures are linked with bonds which help to do the translation without dropping the essence of the text. The experimental research of set of words and gestures was made in the primary classes of special schools for children with hearing impairments.

*Стаття надійшла до редакції 16.06.2011.*